JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **70**, 505–529 (1979)

Vector-Valued Lg-Splines
I. Interpolating Splines

GURSHARAN S. SIDHU

Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, México 20, D.F.

AND

HOWARD L. WEINERT

Department of Electrical Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

Submitted by T. T. Soong

The theory of Lg-splines developed by Jerome and Schumaker is extended to the vector-valued (multivariate) case. The extension is described in the framework of a reproducing-kernel Hilbert space which among other things allows the establishment of a congruent least-squares estimation problem for a vector-valued lumped random process. The results include a dynamic recursive algorithm for vector-valued Lg-splines with EHB data and a useful structural characterization theorem for such splines. Some results on computable approximation error bounds are also included.

I. INTRODUCTION

We shall be concerned in this paper with an extension of the Lg-splines first discussed in a 1969 paper [1] by Jerome and Schumaker. Their theory was developed in the space H_k of functions f, on the interval $W = [0, T]$, whose kth derivative $f^{(k)}$ exists a.e. and is square-integrable on W; i.e.

$$H_k = \left\{ f \text{ on } [0, T] : f^{(k)} \text{ exists a.e. and } \int_0^T \{f^{(k)}\}^2 < \infty \right\}. \tag{1.1}$$

Thus if we define the differential operator $L = D^k + \sum_{j=0}^{k-1} a_j(t)\, D^j$, where $D = d/dt$ and $a_j \in C^j(W)$, it is clear that $f \in H_k$ if and only if $\int_0^T \{Lf\}^2 < \infty$. Now, corresponding to real numbers $\{r_j, 1 \leqslant j \leqslant N\}$ and linear functionals

0022-247X/79/080505-25$02.00/0

$\{\lambda_j, 1 \leqslant j \leqslant N\}$ on H_k Jerome and Schumaker defined an *Lg-spline interpolating* $\{r_j, 1 \leqslant j \leqslant N\}$ *with respect to* $\{\lambda_j, 1 \leqslant j \leqslant N\}$ *as an* $s \in H_k$ *such that*

$$s \in U_r = \{f \in H_k : \lambda_j f = r_j, 1 \leqslant j \leqslant N\} \tag{1.2a}$$

$$\int_0^T (Ls)^2 \leqslant \int_0^T (Lf)^2 \qquad \text{all } f \in U_r. \tag{1.2b}$$

Thus s is the function satisfying the interpolation constraints (1.2a) that is smoothest in the sense of (1.2b). Freedom in choosing L implies a consequent freedom of the choice of the smoothness criterion (1.2b).

Jerome and Schumaker have presented a rather complete development of these splines from a function analytic point of view. Their results include existence and uniqueness conditions (so-called poisedness of (1.2) with respect to L) and an elegant structural theorem. Their uniqueness condition has subsequently been given an operationally useful systems-theoretic interpretation in our work [2].

Lg-splines and their specialized relatives have been found to have a good deal of interesting structure—hence a large body of related theoretical work has been developed. On the other hand, their smoothness and interpolation properties have prompted considerable applications interest with consequent activity on efficient computational algorithms.

From the viewpoint of applications it has become clear that various extensions of Jerome and Schumaker's *Lg*-splines are required:

(A) *Splines on* R^n. These are of interest in various applications, for instance, those related to picture processing ($n = 2$) such as reconstruction from samples.

(B) *Vector-valued* (*Multivariate*) *splines.* Often one wishes to reconstruct a number of functions ($p > 1$) from their sample values. Since the samples of one function can convey information about the others, it is not adequate to interpolate the samples of each function *independently* of those of the other functions. A *simultaneous* interpolation would be more appropriate. This paper is concerned with the resolution of this problem.

(C) ARMA *splines.* It has been noted [3]–[6] that *Lg*-splines provide a solution of the following optimal control problem in $\mathscr{L}_2(W)$:

Determine the minimum energy input function u on $[0, T]$ *such that the output y of the linear system* $Ly = u$ *satisfies the constraints* (1.3)

$$\lambda_j y = r_j, \qquad 1 \leqslant j \leqslant N.$$

Obviously, since $u = Ly$, the optimum solution is $u_* = Ls$ where s is the *Lg*-spline of (1.2). Vector-valued splines would provide an extension of this correspondence to the case of systems with vector-valued inputs and outputs.

However, a severe restriction in (1.3) is the form of the linear systems the model $Ly = u$ corresponds to the so-called ***autoregressive*** (or ***numerator—free***) *systems*. General linear differential systems have two operators L and M associated with them—roughly speaking, they are of the type $Ly = Mu$ and are said to be of *autoregressive moving-average* (ARMA) *type*. This clearly motivates the concept of an ARMA spline. We have already reported such results [5] with related work being given also in [6].

In a subsequent paper we shall discuss the case of vector-valued ARMA splines which brings with it a host of new difficulties.

II. Vector-Valued LG-Splines—Definition, Existence and Uniqueness

For a given integer $p \geqslant 1$, and fixed nonnegative integers $n_1, n_2, \ldots, n_p$, let H be the space of $p \times 1$ vector-valued functions[1] $f = (f_1, f_2, \ldots, f_p)'$ on W such that $f_j \in H_{n_j}$, $1 \leqslant j \leqslant p$; that is to say[2]

$$H = H_{n_1} \times H_{n_2} \times \cdots \times H_{n_p}, \tag{2.1}$$

where each H_{n_j} is defined as in (1.1). Thus for each j, $f_j^{(n_j)} \in \mathscr{L}_2(W)$, the space of square-integrable functions on $[0, T]$.

Let L be a $p \times p$ matrix of ordinary differential operators, with its ijth entry L_{ij} of the form

$$L_{ij} = \sum_{l=0}^{n_j} a_{ij,l}(t)\, D^l, \qquad D = d/dt, \tag{2.2}$$

where we assume (see Appendix 1) that for each j

$$\begin{aligned} a_{ij,n_j} &= 1 \quad \text{if} \quad i = j, \\ &= 0 \quad \text{if} \quad i \neq j. \end{aligned} \tag{2.3}$$

Clearly now

$$f \in H \qquad \text{if and only if} \qquad Lf \in \mathscr{L}_2{}^p(W), \tag{2.4}$$

where $\mathscr{L}_2{}^p(W)$ is the p-fold Cartesian product of $\mathscr{L}_2(W)$.

Now if $\{\lambda_i\}_1^N$ are independent linear functionals on H, then for any real numbers $\{r_j\}_1^N$ one has:

[1] Primes (′) denote matrix transpose.

[2] Note that one could work in the space of functions mapping $\{1, 2, \ldots, p\} \times W$ into the reals. This is in fact implicit in particular in Section V.

DEFINITION 2.1. An $s \in U_r \subset H$ is a vector-valued Lg-spline interpolating $\{r_j, 1 \leqslant j \leqslant N\}$ with respect to $\{\lambda_j, 1 \leqslant j \leqslant N\}$ if

$$\int_0^T (Ls)'(Ls) \leqslant \int_0^T (Lf)'(Lf), \qquad \text{all } f \in U_r \tag{2.5a}$$

where

$$U_r = \{f \in H: \lambda_j f = r_j, 1 \leqslant j \leqslant N\}. \quad \blacksquare \tag{2.5b}$$

H can be made a Hilbert space under a variety of norms (one is given in Section III). First, let us obtain the natural generalization to the vector-valued case of the existence and uniqueness results of [1]:

THEOREM 2.2. (a) *Provided the* $\{\lambda_j, 1 \leqslant j \leqslant N\}$ *are continuous functionals on* H, *the vector-valued spline of definition* 2.1 *always exists.*

(b) *It is unique if and only if either of the following equivalent conditions holds:*

(i) $N_L \cap U_0 = \{0\}$ *where*

$$N_L = \{f \in H: Lf = 0\} \qquad \textit{and} \qquad U_0 = \{f \in H: \lambda_j f = 0, 1 \leqslant j \leqslant N\}; \tag{2.6}$$

(ii) $N \geqslant n = n_1 + n_2 + \cdots + n_p$, *and among the* $\{\lambda_j, 1 \leqslant j \leqslant N\}$ *there are* n *functionals linearly independent on* N_L. (2.7)

Proof. (a) Proceeding as in [1] we can establish that LU_r, the image of U_r under L, is a closed linear variety in $\mathscr{L}_2^p(W)$ and hence contains a minimum $\mathscr{L}_2^p$-norm vector u_*. By the definition of $LU_r = \{Lf: f \in U_r\}$ there must then *exist* an $s \in U_r$ such that $Ls = u_*$. This s is a solution of (2.5).

(b) Let s be a solution of (2.5). Then, so is $s + g$ for any $g \in N_L \cap U_0$. Hence $N_L \cap U_0 = \{0\}$ is a necessary condition for uniqueness. It is also sufficient for, as we show below, the difference $g = s_{(1)} - s_{(2)}$ of any two solutions of (2.5) is in $N_L \cap U_0$. First, since λ_j is linear and $\lambda_j s_{(1)} = \lambda_j s_{(2)} = r_j$ we have $s_{(1)} - s_{(2)} \in U_0$. Also, from the proof of part (a), $Ls_{(1)} = Ls_{(2)} = u_*$ and thence $s_{(1)} - s_{(2)} \in N_L$.

To establish (ii) note that $u_* \in LU_r$ is unique. Hence the determination of s involves solving the system $Ls = u_*$ of p differential equations subject to boundary conditions $\lambda_j s = r_j$, $1 \leqslant j \leqslant N$. From the discussion of appendix I it is clear that the solution of these *de*'s is unique if and only if $n = n_1 + n_2 + \cdots + n_p$ of the λ_j are linearly independent on N_L. $\blacksquare$

Henceforth we shall assume the conditions for existence and uniqueness, namely that the λ_j are continuous on H, that $N \geqslant n$, and that, after possible rearrangement,

$$\{\lambda_j, 1 \leqslant j \leqslant n\} \text{ are linearly independent on } N_L. \tag{2.8}$$

III. H as a Hilbert Space

The integer

$$n = \sum_{j=1}^{p} n_j \tag{3.1}$$

has a special significance:

LEMMA 3.1. *N_L is a subspace of H with dimension equal to n.* ∎

This is proved in Appendix I using a state-space representation of $Lf = 0$.

As a consequence of Lemma 3.1, N_L has a basis $\{z_j, 1 \leqslant j \leqslant n\}$ chosen to be dual to the $\{\lambda_j, 1 \leqslant j \leqslant n\}$. Thus for $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant n$

$$Lz_j = 0, \qquad \lambda_i z_j = \delta_{ij} = 1 \quad \text{for } i = j \\ = 0 \quad \text{for } i \neq j. \tag{3.2}$$

Next, let $G(\cdot, \cdot)$ be the Green's function of L relative to $\{\lambda_j, 1 \leqslant j \leqslant n\}$, i.e. G is the $p \times p$ matrix of functions on $W \times W$ given, for each $t \in W$, by

$$LG(\cdot, t) = I_p \delta(\cdot - t), \qquad \lambda_j G(\cdot, t) = \mathbf{0}', \qquad 1 \leqslant j \leqslant n \tag{3.3}$$

where I_p is the $p \times p$ identity matrix, and $\mathbf{0}$ a $p \times 1$ column of zeros. In (3.3) λ_j operates on the columns of $G(\cdot, t)$.

Now, any $f \in H$ can be written in the form

$$f(\cdot) = \sum_{j=1}^{n} (\lambda_j f)\, z_j(\cdot) + \int_0^T G(\cdot, t)\, \{Lf(t)\}\, dt \tag{3.4}$$

which prompts the inner product for H

$$\langle f, g \rangle_H = \sum_{j=1}^{n} (\lambda_j f)\,(\lambda_j g) + \int_0^T \{Lf(t)\}' \{Lg(t)\}\, dt \tag{3.5}$$

with the corresponding norm

$$\| f \|_H^2 = \sum_{j=1}^{n} (\lambda_j f)^2 + \int_0^T \{Lf(t)\}' \{Lf(t)\}\, dt. \tag{3.6}$$

Note that the first term in (3.4) belongs to N_L and is orthogonal to the second term. Thus (3.4) implies an orthogonal decomposition of H:

$$H = N_L \oplus H_1, \qquad N_L \perp H_1 \tag{3.7}$$

It is easily seen that N_L and H_1 are congruent to $\mathbb{R}^n$ and $\mathscr{L}_2{}^p(W)$ respectively.

Thus, since $\mathbb{R}^n$ and $\mathscr{L}_2{}^p(W)$ are complete so are N_L and H_1. We can then assert the following for their direct sum H:

LEMMA 3.2. *H equipped with the inner product* (3.5) *is a Hilbert space.*

Note that $\sum_{j=1}^{n} (\lambda_j f)^2$ is constant on U_r; thus definition 2.1 can be put in the minimum-norm form:

THEOREM 3.3. *The multivariate Lg-spline s interpolating $\{r_j, 1 \leqslant j \leqslant N\}$ is the unique minimum norm element of the linear variety U_r of H.* ∎

As a consequence, one can use the projection theorem [7, p. 64] to establish that

$$s = \mathscr{P}^{\mathscr{S}_N} g \qquad \text{any} \qquad g \in U_r, \tag{3.9}$$

where $\mathscr{P}^{\mathscr{S}_N}$ denotes projection onto $\mathscr{S}_N$, the orthogonal complement of U_0. It is easy to show that

$$\mathscr{S}_N = U_0{}^{\perp} = \operatorname{span}\{h_j, 1 \leqslant j \leqslant N\}, \tag{3.10}$$

where h_j are the representers of the λ_j; i.e.

$$h_j \in H \qquad \text{is such that} \qquad \lambda_j f = \langle f, h_j \rangle_H \quad \text{all} \quad f \in H. \tag{3.11}$$

Solving the normal equations corresponding to (3.9) one has

$$s(\cdot) = \mathbf{h}'(\cdot)\, \mathbf{R}^{-1} \mathbf{r} \tag{3.12}$$

where $\mathbf{h}'(\cdot) = (h_1(\cdot), h_2(\cdot),..., h_N(\cdot))$, $\mathbf{r}' = (r_1, r_2,..., r_N)$, and $\mathbf{R}$ is the $N \times N$ matrix with ijth entry $\langle h_i, h_j \rangle_H$. The solution (3.12) is in fact not very useful in practice and is include here primarily for completeness.

IV. REPRODUCING KERNEL FOR H

Let K be the $p \times p$ matrix-valued function on $W \times W$ defined by

$$K(t, \tau) = \sum_{j=1}^{n} z_j(t)\, z_j'(\tau) + \int_0^T G(t, \xi)\, G'(\tau, \xi)\, d\xi \tag{4.1}$$

and let $K_j{}^r$ and $K_l{}^c$ denote respectively the jth row and lth column of K. The function $K(\cdot, \cdot)$ is of central importance in this paper. We start by establishing some of its key properties. Proofs are given only for the nonobvious ones.

PROPERTY 4.1. *Symmetry:*

$$K(t, \tau) = K'(\tau, t), \qquad all \qquad t, \tau \in W. \tag{4.2a}$$

Thus

$$K_j^c(t, \tau) = [K_j^r(\tau, t)]'. \tag{4.2b}$$

PROPERTY 4.2. *Inclusion of K in H:*

From (4.1) and (3.4) is is evident that, for each *fixed* $t \in W$,

$$K_j^c(\cdot, t) = [K_j^r(t, \cdot)]' \in H, \qquad 1 \leqslant j \leqslant p, \tag{4.3a}$$

Thus for the ijth entry K_{ij} of K we have

$$K_{ij}(\cdot, t) \in H_{n_i}; \qquad K_{ij}(t, \cdot) \in H_{n_j}. \tag{4.3b}$$

PROPERTY 4.3. *Reproducing Property:*

From (4.1), (3.2)–(3.3), *for each* $t \in W$,

$$LK(\cdot, t) = G'(t, \cdot), \tag{4.4a}$$

$$\lambda_j K(\cdot, t) = z_j'(t), \qquad 1 \leqslant j \leqslant n. \tag{4.4b}$$

Hence, using (3.5) it is readily verified that, for every $f = (f_1, f_2, ..., f_p)' \in H$, and *for each* $t \in W$,

$$\langle K_j^c(\cdot, t), f(\cdot)\rangle_H = \langle K_j^r(t, \cdot), f(\cdot)\rangle_H = f_j(t), \qquad 1 \leqslant j \leqslant p. \tag{4.5}$$

This is known as the *reproducing property* [8] *of* K.

H is said to be a *reproducing kernel Hilbert space* (RKHS) with K as its *reproducing kernel* (RK).

PROPERTY 4.4. *Nonnegative Definiteness of K:*

Using (4.5) it is easily shown that for any finite integer $m \geqslant 1$ and for each choice of $\tau_1, \tau_2, ..., \tau_m$ in W and real $p \times 1$ vectors $a_1, a_2, ..., a_m$:

$$\sum_{i=1}^{m} \sum_{j=1}^{m} a_i' K(\tau_i, \tau_j)\, a_j = \left\| \sum_{i=1}^{m} K(\cdot, \tau_i)\, a_i \right\|_H^2 \geqslant 0. \tag{4.6}$$

Thus K is a matrix-valued nonnegative definite function on $W \times W$ in the sense of Mercer [9].

PROPERTY 4.5. *The* RK *and the Representers of Continuous Linear Functionals:*

The representer h_i of λ_i is given by

$$h_i'(t) = \lambda_i K(\cdot, t) = [\lambda_i K_1^c(\cdot, t), ..., \lambda_i K_p^c(\cdot, t)], \tag{4.7}$$

and

$$\langle h_i, h_j\rangle_H = \lambda_{j(\tau)}[\lambda_{i(t)}K(t, \tau)]'. \tag{4.8}$$

Here the parenthetical subscripts on λ_i and λ_j show the independent variables with respect to which they operate.

Proof of (4.7)–(4.8). Any $f = (f_1, f_2, ..., f_p)' \in H$ can be decomposed in H in the form $f = \sum_{j=1}^{p} f_{[j]}$ where $f_{[j]} = (0, 0, ..., 0, f_j, 0, ..., 0)'$.

The linearity of λ_i then implies

$$\lambda_i f = \sum_{j=1}^{p} \lambda_i f_{[j]}. \tag{4.9}$$

Clearly thus λ_i induces continuous linear functionals $\lambda_{ij}: H_{n_j} \to \mathbb{R}$, and (4.9) can be written as:

$$\lambda_i f = \sum_{j=1}^{p} \lambda_{ij} f_j. \tag{4.10}$$

Thus using (4.5)

$$\begin{aligned}\lambda_{i(t)} f(t) &= \sum_{j=1}^{p} \lambda_{ij(t)} \langle K_j^c(\cdot, t), f(\cdot)\rangle_H \\ &= \left\langle \sum_{j=1}^{p} \lambda_{ij(t)} K_j^c(\cdot, t), f(\cdot)\right\rangle_H\end{aligned} \tag{4.11}$$

which implies that (note (4.2))

$$\begin{aligned}h_i'(\cdot) &= \left[\sum_{j=1}^{p} \lambda_{ij(t)} K_{1j}(\cdot, t), ..., \sum_{j=1}^{p} \lambda_{ij(t)} K_{pj}(\cdot, \tau)\right] \\ &= \left[\sum_{j=1}^{p} \lambda_{ij(t)} K_{j1}(t, \cdot), ..., \sum_{j=1}^{p} \lambda_{ij(t)} K_{jp}(t, \cdot)\right].\end{aligned}$$

In view of (4.10) this gives us (4.7). Equation (4.8) is a ready consequence of (4.7) for, $\langle h_i, h_j\rangle_H = \lambda_j h_i(\cdot) = \lambda_j[\lambda_{i(\tau)} K(\tau, \cdot)]'$.

For future use we note that (4.7), (4.4b) and (3.2) imply

$$z_j = h_j, \qquad \langle h_i, h_j\rangle_H = \delta_{ij}, \qquad 1 \leqslant i, \quad j \leqslant n. \tag{4.12}$$

PROPERTY 4.6. *The* RK *spans* H:

As a consequence of (4.5), $\langle f(\cdot), K_j^c(\cdot, t)\rangle_H = 0$ for each j and every $t \in W$ if and only if $f = 0$. Thus we have

$$\begin{aligned}H &= \operatorname{span}\{K_j^c(\cdot, t), 1 \leqslant j \leqslant p, t \in W\} \\ &= \operatorname{span}\{K_j^r(t, \cdot), 1 \leqslant j \leqslant p, t \in W\}.\end{aligned} \tag{4.13}$$

It can be shown furthermore that the RK is unique. Thus clearly K completely determines H.

The following theorem is of central importance:

THEOREM 4.1. *For each* $t \in W$

$$s'(t) = \mathscr{V}\mathscr{P}^{\mathscr{S}_N} K(\cdot, t), \tag{4.14}$$

where $\mathscr{V}: \mathscr{S}_N \to \mathbb{R}$ *through* $\mathscr{V}h_j = r_j$, $1 \leqslant j \leqslant N$, *and* $\mathscr{V}$ *linear.*

Proof. Let $s(\cdot) = [s_1(\cdot), s_2(\cdot),..., s_p(\cdot)]'$. Using (4.5) and the selfadjointness of $\mathscr{P}^{\mathscr{S}_N}$ (3.9) can be rewritten

$$s_j(t) = \langle \mathscr{P}^{\mathscr{S}_N} g(\cdot), K_j^c(\cdot, t)\rangle_H = \langle g(\cdot), \mathscr{P}^{\mathscr{S}_N} K_j^c(\cdot, t)\rangle_H$$

which upon expressing $\mathscr{P}^{\mathscr{S}_N} K_j^c(\cdot, t) = \sum_{i=1}^N \beta_{ji}(t)\, h_i(\cdot)$ gives

$$s_j(t) = \sum_{i=1}^N \beta_{ji}(t)\, \langle g(\cdot), h_i(\cdot)\rangle_H = \sum_{i=1}^N \beta_{ji}(t)\, r_i$$

since $g \in U_r$. ∎

One could go directly from Theorem 4.1 to the algorithm of Section VI using an extension of our approach of [10]. Instead we choose to proceed via the intermediate step of introducing a space Y of random variables that is isometrically isomorphic to H, and then obtain the algorithm in Y. This is done since the type of algorithm obtained is much better known in the related context of linear least-squares estimation for lumped random processes.

V. Splines and Stochastic Estimation

From properties 4.2 and 4.4, K is a symmetric nonnegative definite function on $W \times W$. Thus there exists [11] a zero-mean $p \times 1$ vector-valued process $\{y(t), t \in W\}$ with covariance function K:

$$y(t) = (y_1(t), y_2(t),..., y_p(t))'; \qquad E\{y_j(t)\} = 0 \tag{5.1}$$

$$E\{y(t)\, y'(\tau)\} = K(t, \tau); \qquad t, \tau \in W. \tag{5.2}$$

Ley Y be the complete vector space of finite variance, zero-mean random variables spanned by the random variables $\{y_j(t); 1 \leqslant j \leqslant p, t \in W\}$ and with the usual inner product $\langle a, b\rangle_Y = E\{ab\}$. We use E to denote probabilistic expectation with respect to the probability law of the process $y(\cdot)$.

From (4.3), (4.5) and (5.2) it is clear that

$$\langle K_i^r(t, \cdot), K_j^r(\tau, \cdot)\rangle_H = K_{ij}(t, \tau) = \langle y_i(t), y_j(\tau)\rangle_Y \tag{5.3}$$

Since $\{K_i^r(t, \cdot); 1 \leqslant i \leqslant p, t \in W\}$ and $\{y_i(t); 1 \leqslant i \leqslant p, t \in W\}$ respectively span H and Y, (5.3) implies the following extension of a theorem first stated by Loeve [12, p. 408]:

THEOREM 5.1. *The space H with* RK $K(\cdot, \cdot)$ *and the space Y of random variables defined above are isometrically isomorphic (congruent). Furthermore $f \in H$ corresponds to $z \in Y$, denoted $f \sim z$, if and only if*

$$f_j(t) = E\{zy_j(t)\}. \quad \blacksquare \tag{5.4}$$

It is readily noted that under this congruence

$$K_j^c(\cdot, t) \quad \text{or} \quad K_j^r(t, \cdot) \sim y_j(t), \qquad 1 \leqslant j \leqslant p, \tag{5.5}$$

$$h_j(\cdot) = [\lambda_{j(t)} K(t, \cdot)]' \sim \lambda_{j(t)} y(t), \qquad 1 \leqslant j \leqslant N. \tag{5.6}$$

Thus if $\mathscr{D}_N = \operatorname{span}\{\lambda_j y, 1 \leqslant j \leqslant N\}$ then

$$\mathscr{S}_N \sim \mathscr{D}_N . \tag{5.7}$$

Consequently

$$\mathscr{P}^{\mathscr{S}_N} K(t, \cdot) \sim \mathscr{P}^{\mathscr{S}_N} y(t), \qquad \text{each} \qquad t \in W \tag{5.8}$$

where $\mathscr{P}^{\mathscr{D}_N} y(t) = (\mathscr{P}^{\mathscr{D}_N} y_1(t), \mathscr{P}^{\mathscr{D}_N} y_2(t), \ldots, \mathscr{P}^{\mathscr{D}_N} y_p(t))$. The random vector $\mathscr{P}^{\mathscr{D}_N} y(t)$ is known as *the linear least-squares estimate (llse) of $y(t)$ given* $\{\lambda_j y, 1 \leqslant j \leqslant N\}$ since it minimizes, for $1 \leqslant j \leqslant p$, $E(y_j(t) - \rho_j)^2$ over all $\rho_j \in \mathscr{D}_N$. If we denote by $\hat{y}(t)$ the sample value of $\mathscr{P}^{\mathscr{D}_N} y(t)$ corresponding to $\{\lambda_j y = r_j, 1 \leqslant j \leqslant N\}$ then we have the following generalization of the univariate results of [13]–[17].

THEOREM 5.2.

$$s(t) = \hat{y}(t). \quad \blacksquare \tag{5.9}$$

It is natural thus to seek algorithms for computing $\hat{y}(\cdot)$ and hence $s(\cdot)$. This is greatly facilitated by the fact that $y(\cdot)$ is a lumped process—it has a known finite-dimensional linear model as summarized in the following theorem:

THEOREM 5.3. *The random process $y(\cdot)$ of* (5.1) *is such that*

$$Ly = u \tag{5.10}$$

where $u(\cdot)$ is a $p \times 1$ vector-valued, zero mean, white process, i.e.,

$$Eu(t) \equiv 0; \qquad Eu(t)\, u'(\tau) = I_p \delta(t - \tau). \quad \blacksquare \tag{5.11}$$

The proof is quite easy to obtain, albeit somewhat informally, since $y_j(t) \sim K_j^r(t, \cdot)$ implies that $L_{(t)}K(t, \cdot) = G'(\cdot, t) \sim L_{(t)} y(t)$. Writing $L_{(t)} y(t) = (u_1(t), u_2(t),..., u_p(t))'$ one has $u_j(t) \sim G_j^c(\cdot, t)$, and hence from (3.5) it is readily verified that

$$Eu_i(t)\, u_j(\tau) = \langle G_i^c(\cdot, t), G_j^c(\cdot, \tau)\rangle_H = \delta_{ij}\delta(t - \tau).$$

Now, in view of Appendix I we have a ready corollary that $y(\cdot)$ has a finite-dimensional state model:

COROLLARY 5.4. *The process $y(\cdot)$ is the output of the state model*

$$y(t) = Cx(t); \qquad \frac{d}{dt} x(t) = A(t)\, x(t) + Bu(t) \tag{5.12a}$$

where $u(\cdot)$ is as in (5.11) *and $A(\cdot)$, B, C are $n \times n$, $n \times p$ and $p \times n$ matrices given by*

$$B = \text{block diag}[[0,..., 0, 1]', n_j \times 1] \tag{5.12b}$$

$$C = \text{block diag}[[1, 0,..., 0], 1 \times n_j] \tag{5.12c}$$

$$A(\cdot) = \text{block}[A_{ij}, n_i \times n_j] \tag{5.12d}$$

$$A_{ii}(\cdot) = \left[\begin{array}{c|ccc} \mathbf{0} & & I_{n_i-1} & \\ \hline -a_{ii,0} & -a_{ii,1} & \cdots & -a_{ii,n_i-1} \end{array}\right], \qquad n_i \times n_i, \tag{5.12e}$$

$$A_{ij}(\cdot) = \left[\begin{array}{cccc} & & \mathbf{0} & \\ \hline -a_{ij,0} & -a_{ij,1} & \cdots & -a_{ij,n_j-1} \end{array}\right], \qquad n_i \times n_j, \quad i \neq j \tag{5.12f}$$

and $x(\cdot)$ is $n \times 1$ given by

$$x' = [y_1, y_1^{(1)},..., y^{(n_1-1)} \mid y_2, y_2^{(1)},..., y^{(n_2-1)} \mid \cdots \mid y_p, y_p^{(1)},..., y_p^{(n_p-1)}]. \tag{5.12g}$$

The model in (5.12) is not complete in that it does not allow us to completely obtain the covariance function of $y(\cdot)$. What is lacking is a set of boundary conditions for the differential system (5.12a). Although we can go further without additional restriction, the most fruitful results accrue for a very broad special class of λ_j considered in the next section.

VI. LG-SPLINES WITH EHB DATA

We now restrict attention to λ_j, $1 \leqslant j \leqslant N$, of the form

$$\lambda_j f = \sum_{i=1}^{p} \sum_{l=1}^{n_i} \alpha_{ij,l} f_i^{(l-1)}(t_j), \qquad t_j \in W, \tag{6.1}$$

where t_j and $\alpha_{ij,l}$, $1 \leqslant i \leqslant p$, $1 \leqslant l \leqslant n_i$ are known real numbers. Such functionals are a natural generalization to our setting of the so-called *extended Hermite–Birkhoff* (EHB) *functionals* [1]. With this restriction we shall refer to s as an *Lg-spline with* EHB *data.*

We assume that the *knots* t_j are ordered thus:

$$0 \leqslant t_1 \leqslant t_2 \leqslant \cdots \leqslant t_N \leqslant T. \tag{6.2}$$

If we define $1 \times n$ matrices, for $1 \leqslant j \leqslant N$,

$$c_j = (c_{j1}, c_{j2}, \ldots, c_{jp}); \qquad c_{ji} = (\alpha_{ji,1}, \alpha_{ji,2}, \ldots, \alpha_{ji,n_i}), \tag{6.3}$$

then

$$\lambda_j f = \sum_{j=1}^{p} c_{ji}(f_i(t_j), f_i^{(1)}(t_j), \ldots, f_i^{(n_i-1)}(t_j))'. \tag{6.4}$$

Thus in view of (5.12g) the random variables $\lambda_j y$ are given by

$$\lambda_j y = c_j x(t_j), \tag{6.5}$$

where $x(\cdot)$ is the random vector defined in (5.12). ∎

Now the missing boundary conditions of (5.12) can be obtained:

THEOREM 6.1. *When the λ_j are of* EHB *type the boundary conditions of model* (5.12) *are given by:*

$$x(t_n) = \mathcal{O}^{-1}\left\{\Lambda + \int_{t_1}^{t_n} \Delta(\xi)\, Bu(\xi)\, d\xi\right\}, \tag{6.6}$$

$$\Pi_n = E\{x(t_n)\, x'(t_n)\} = \mathcal{O}^{-1}\{I_n + Q\}\, \mathcal{O}^{-T}, \tag{6.7}$$

$$\begin{aligned} E\{x(t_n)\, u'(\xi)\} &= 0, && \xi < t_1 \quad \text{or} \quad \xi > t_n, \\ &= \mathcal{O}^{-1}\Delta(\xi)\, B, && t_1 \leqslant \xi \leqslant t_n, \end{aligned} \tag{6.8}$$

where

$$\Lambda' = (\lambda_1 y, \lambda_2 y, \ldots, \lambda_n y), \tag{6.9}$$

$$\mathcal{O} = \begin{bmatrix} c_1\phi(t_1, t_n) \\ \vdots \\ c_{n-1}\phi(t_{n-1}, t_n) \\ c_n \end{bmatrix}, \qquad n \times n \text{ matrix}, \tag{6.10}$$

$$Q = \int_{t_1}^{t_n} \Delta(\xi)\, BB'\Delta'(\xi)\, d\xi, \qquad n \times n \text{ matrix}, \tag{6.11}$$

and $\Delta(\xi)$ is the $n \times n$ matrix with ith *row*

$$\begin{aligned} \Delta_i(\xi) &= c_i\phi(t_i, \xi), && t_i \leqslant \xi \leqslant t_n, \\ &= 0, && \text{otherwise}. \end{aligned} \tag{6.12}$$

Here ϕ is the state transition matrix of $A(\cdot)$, i.e.

$$\frac{\partial}{\partial t}\phi(t,\tau) = A(t)\,\phi(t,\tau), \qquad \phi(\tau,\tau) = I_n. \tag{6.13}$$

Proof. First, $y_j(t) \sim K_j{}^r(t,\cdot)$ implies that $z_i(\cdot) \sim \lambda_i y$, $1 \leqslant i \leqslant n$ (note (4.12), (5.6)). Thus (4.2) and Theorem 5.1 imply that for $1 \leqslant i, j \leqslant n$, $E\{(\lambda_i y)(\lambda_j y)\} = \delta_{ij}$ and hence

$$E\Lambda\Lambda' = I_n. \tag{6.14}$$

Also, $N_L \perp H_1$ implies that $\int_0^T G(\cdot,\xi) f(\xi)\,d\xi \perp z_j(\cdot)$, $1 \leqslant j \leqslant n$, for any $f \in \mathscr{L}_2{}^p(W)$. Thus in the congruent space Y we must have

$$\int_0^T f'(\xi)\,u(\xi)\,d\xi \perp \lambda_j y, \qquad 1 \leqslant j \leqslant n. \tag{6.15}$$

Using the definition (6.13) of ϕ it is easily seen that (5.12) has an integral form:

$$y(t) = Cx(t); \qquad x(t) = \phi(t,t_n)\,x(t_n) + \int_{t_n}^{t} \phi(t,\xi)\,Bu(\xi)\,d\xi. \tag{6.16}$$

Hence using (6.5)

$$\Lambda = \mathcal{O}x(t_n) - \int_{t_1}^{t_n} \Delta(\xi)\,Bu(\xi)\,d\xi, \tag{6.17}$$

of which (6.6) is an obvious consequence provided $\mathcal{O}$ is nonsingular. It is easily shown by using our technique of [2] that $\mathcal{O}$ is nonsingular if and only if condition (2.8) holds.

Now, in view of (6.4)–(6.15) we obtain from (6.17)

$$\mathcal{O}\Pi_n\mathcal{O}' = I + Q, \tag{6.18}$$

$$\begin{aligned} \mathcal{O}E\{x(t_n)\,u'(\xi)\} &= 0 + \int_{t_1}^{t_n} \Delta(\tau)\,BE\{u(\tau)\,u'(\xi)\}\,d\tau, \\ &= 0, \qquad \text{if} \qquad \xi \notin [t_1, t_n] \\ &= \Delta(\xi)\,B, \qquad \text{if} \qquad t_1 \leqslant \xi \leqslant t_n. \end{aligned} \tag{6.19}$$

Now (6.7)–(6.8) follow from (6.18)–(6.19). ∎

Having obtained a complete model for $\{y(t), t \in W\}$ we can now give a recursive algorithm for computing the sample function $\hat{y}(t)$ of its linear least-squares estimate $\mathscr{P}^{\mathscr{D}_N}y(t)$ given observations $\{\lambda_j y, 1 \leqslant j \leqslant N\}$. This, in view of Theorem 5.2 is also an algorithm for $s(\cdot)$.

VI.1. *Dynamical Recursive Algorithm*

$$s(t) = C\hat{x}(t \mid N), \qquad t \in W \tag{6.20}$$

where the $n \times 1$ vector function $\hat{x}(\cdot \mid N)$ is computed through the following recursive procedure consisting of algebraic updates at knots and differential equations (with determined end conditions) between adjacent knots.

Step 1. Initialization

(i) Compute Q_{j+1} and $M_{j+1}(t_{j+1})$ recursively for $1 \leqslant j \leqslant n-1$ through

$$M_1(t_1) = c_1; \qquad Q_1 = \mathbf{0} \quad (n \times n \text{ matrix}); \tag{6.21}$$

$$\Pi_j(t_j) = \mathbf{0} \qquad (n \times n \text{ matrix}), \qquad 1 \leqslant j \leqslant n-1 \tag{6.22}$$

$$\frac{d}{dt} M_j(t) = -M_j(t)\, A(t), \tag{6.23}$$

$$t_j \leqslant t \leqslant t_{j+1}$$

$$\frac{d}{dt} \Pi_j(t) = A(t)\, \Pi_j(t) + \Pi_j(t)\, A'(t) + BB', \tag{6.24}$$

$$M'_{j+1}(t_{j+1}) = [M'_j(t_{j+1}) \mid c'_{j+1}) \tag{6.25}$$

$$Q_{j+1} = Q_j + \left[\begin{array}{c|c} M_j(t_{j+1}) \cdot \Pi_j(t_{j+1}) \cdot M'_j(t_{j+1}) & \mathbf{0}' \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \tag{6.26}$$

(ii) Let $\mathcal{O} = M_n(t_n)$, $Q = Q_n$, and compute

$$x_0 = \mathcal{O}^{-1}(r_1, r_2, \ldots, r_n)'; \qquad \Pi_0 = \mathcal{O}^{-1} Q \mathcal{O}^{-T} \tag{6.27}$$

Step 2. Forward pass for $t \geqslant t_n$:

(iii) Set

$$\hat{x}(t_n \mid n) = x_0; \qquad P(t_n \mid n) = \Pi_0 . \tag{6.28}$$

(iv) Recursively compute and store $\{e_{j+1}, R^{\epsilon}_{j+1}, K'_{j+1}, n \leqslant j \leqslant N-1\}$ through

$$e_{j+1} = r_{j+1} - c_{j+1}\hat{x}(t_{j+1} \mid j); \qquad R^{\epsilon}_{j+1} = c_{j+1} P(t_{j+1} \mid j)\, c'_{j+1}, \tag{6.29}$$

$$K'_{j+1} = P(t_{j+1} \mid j)\, c'_{j+1}; \qquad (n \times 1 \text{ vector}), \tag{6.30}$$

where $\hat{x}(t_{j+1} \mid j)$, $P(t_{j+1} \mid j)$ are computed for $n \leqslant j \leqslant N-1$ through:

$$\frac{d}{dt} \hat{x}(t \mid j) = A(t)\, \hat{x}(t \mid j) \tag{6.31}$$

$$t_j \leqslant t \leqslant t_{j+1}$$

$$\frac{d}{dt} P(t \mid j) = A(t)\, P(t \mid j) + P(t \mid j)\, A'(t) + BB' \tag{6.32}$$

$$\hat{x}(t_{j+1} \mid j+1) = \hat{x}(t_{j+1} \mid j) + K_{j+1}(R_{j+1}^{\epsilon})^{-1} e_{j+1} \tag{6.33}$$

$$P(t_{j+1} \mid j+1) = P(t_{j+1} \mid j) - K_{j+1}(R_{j+1}^{\epsilon})^{-1} K_{j+1}' \tag{6.34}$$

(v) Compute and store $\hat{x}(t_N \mid N)$ using (6.33).

Step 3. Computation of $\hat{x}(t \mid N)$:

(vi) Compute $\hat{x}(t \mid N)$ by integrating, starting at $t = t_N$ with the value of $\hat{x}(t_N \mid N)$ from step 2(v), the equations

$$\frac{d}{dt} \hat{x}(t \mid N) = A(t)\, \hat{x}(t \mid N) + BB'\mu(t \mid N) \tag{6.35}$$

where $\mu(t \mid N)$ is a piecewise continuous $n \times 1$ vector-valued function given by:

$$\begin{aligned} \mu(t \mid N) &= 0, \qquad && t < t_1 \quad \text{or} \quad t > t_N \\ &= \mu_j(t \mid N), && t_{j-1} < t < t_j, \quad 2 \leqslant j \leqslant N. \end{aligned} \tag{6.36}$$

(vii) The pieces $\mu_j(\cdot \mid N)$ are computed recursively, for $2 \leqslant j \leqslant N$, through

$$\mu_{N+1}(t_N \mid N) = 0 \tag{6.37}$$

$$\begin{aligned} \mu_j(t_j \mid N) &= \mu_{j+1}(t_j \mid N) + c_j'(R_j^{\epsilon})^{-1} \{e_j - K_j'\mu_{j+1}(t_j \mid N)\}, \qquad && j > n \\ &= \mu_{j+1}(t_j \mid N) - c_j'\beta_j, && j \leqslant n \end{aligned} \tag{6.38}$$

where

$$(\beta_1, \beta_2, \ldots, \beta_n)' = \mathcal{O}^{-T}\mu_{n+1}(t_n \mid N); \tag{6.39}$$

$$-\frac{d}{dt} \mu_j(t \mid N) = A'(t)\, \mu_j(t \mid N), \qquad t_{j-1} \leqslant t \leqslant t_j. \tag{6.40}$$

Remark 6.1. *Coincident Knots:*

If knots are coincident, i.e. $t_j = t_{j+1}$, then the integration of differential equations between t_j and t_{j+1} is trivialized; to wit, equations (6.23), (6.24), (6.31), (6.32), (6.41) are replaced by the identities:

$$M_j(t_{j+1}) = M_j(t_j), \tag{6.23$'$}$$

$$\Pi_j(t_{j+1}) = \Pi_j(t_{j+1}), \tag{6.24$'$}$$

$$\hat{x}(t_{j+1} \mid j) = \hat{x}(t_j \mid j), \tag{6.31$'$}$$

$$P(t_{j+1} \mid j) = P(t_j \mid j), \tag{6.32$'$}$$

$$\mu_j(t_{j-1} \mid N) = \mu_j(t_j \mid N). \tag{6.41$'$}$$

Remark 6.2. Equation (6.34) has an alternate form:

$$P(t_{j+1} \mid j+1) = \{I - K_{j+1}(R^{\epsilon}_{j+1})^{-1} c'_{j+1}\} P(t_{j+1} \mid j) \{I - K_{j+1}(R^{\epsilon}_{j+1})^{-1} c'_{j+1}\}'. \tag{6.34'}$$

The choice between (6.34) and (6.34′) is a tradeoff between the lower number of computations needed in (6.34) and the better numerical properties of (6.34′). For further remarks on this aspect see [10].

The proof of the algorithm is postponed till Section VIII; let us instead examine a structural theorem that follows from the algorithm.

VII. A Structural Theorem for the EHB-Data Case

In view of Theorem 5.2 it is clear that $s(\cdot)$ is an Lg-spline interpolating $\{r_j\}_1^N$ with respect to $\{\lambda_j\}_1^N$ *if and only if* $s(\cdot) = \hat{y}(\cdot)$, the sample function of the l.l.s.e. of $y(\cdot)$ corresponding to observations $\lambda_j y = r_j$, $1 \leqslant j \leqslant N$. Now in turn the construction of the previous section shows that $\hat{y}(\cdot)$ is this l.l.s.e. if and only if it is the solution of the algorithm (6.20–6.40). This argument is central to establishing the structural theorem of this section.

First, from (6.35), (6.20) and (5.12b–g) it is immediate that if $s = (s_1, s_2, ..., s_p)'$ then

$$\hat{x}(\cdot \mid N) = (s_1, s_1^{(1)}, ..., s_1^{(n_1-1)} \mid \cdots \mid s_p, s_p^{(1)}, ..., s_p^{(n_p-1)})' \tag{7.1}$$

and that

$$Ls(t) = w(t) = B'\mu(t \mid N). \tag{7.2}$$

The $p \times 1$ vector-valued function $w(\cdot)$ in turn has the state model (use (6.40)):

$$w(t) = B'\mu(t \mid N), \qquad -\frac{d}{dt}\mu(t \mid N) = A'(t)\,\mu(t \mid N); \qquad t_{j-1} < t < t_j. \tag{7.3}$$

Now, using the defining relations (5.12) for $A(\cdot)$ and B, it is quite easy to show that (7.3) is equivalent to

$$L^* w(t) = 0, \qquad t_{j-1} < t < t_j, \quad 2 \leqslant j \leqslant N, \tag{7.4}$$

where L^* is a $p \times p$ matrix whose ijth entry L^*_{ij} is the ordinary differential operator given by:

$$L^*_{ij} f = \sum_{k=0}^{n_j} (-1)^k D^k\{a_{ji,k} f\}, \qquad 1 \leqslant i, \quad j \leqslant p, \tag{7.5}$$

Furthermore, straightforward calculations show that (6.40) implies that

$$\mu(t \mid N) = (\Theta_{1,0}s(t), \Theta_{1,1}s(t),..., \Theta_{1,n_1-1}s(t),..., \Theta_{p,0}s(t), \Theta_{p,1}s(t),..., \Theta_{p,n_p-1}s(t))' \tag{7.6}$$

where $\Theta_{i,j}$, are $1 \times p$ rows of ordinary differential operators defined for $1 \leqslant i \leqslant p$, $0 \leqslant j \leqslant n_i - 1$ by

$$\Theta_{i,j}s(t) = [\theta_{i,j}(1), \theta_{i,j}(2),..., \theta_{i,j}(p)]\, Ls(t), \tag{7.7a}$$

$$\theta_{i,j}(k) f = \sum_{l=0}^{n_i-1-j} (-1)^l D^l\{a_{ki,j+1+l} f\}. \tag{7.7b}$$

A knot $\tau = t_{j+1}$ shall be said to have *order* $l(\tau)$ if

$$t_j < t_{j+1} = t_{j+2} = \cdots = t_{j+l(\tau)} < t_{j+l(\tau)+1}; \qquad \tau = t_{j+1}\,. \tag{7.8}$$

Let

$$[f]_\tau = f(\tau+) - f(\tau-). \tag{7.9}$$

Then from (6.38) we have:

$$[\mu(\tau \mid N)]_\tau = \begin{bmatrix} [\Theta_{1,0}s]_\tau \\ \vdots \\ [\Theta_{p,n_p-1}s]_\tau \end{bmatrix} = [c'_{j+1}, c'_{j+2},..., c'_{j+l(\tau)}] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{l(\tau)} \end{bmatrix} = \mathbf{M}\boldsymbol{\alpha} \tag{7.10}$$

for suitable real numbers α_j.

The linear independence of $\{\lambda_k, j+1 \leqslant k \leqslant j + l(\tau)\}$ implies that $l(\tau) \leqslant n$ and that the $n \times l(\tau)$ matrix $\mathbf{M}$ has full rank. Thus $\mathbf{M}$ can be extended to a full rank $n \times n$ matrix $\mathbf{M}_+ = [\mathbf{M} \mid \Gamma]$, which together with the $n \times 1$ vector $\boldsymbol{\alpha}_+ = (\boldsymbol{\alpha}' \mid \mathbf{0})'$ allows us to rewrite (7.10) as

$$[\mu(\tau \mid N)]_\tau = \mathbf{M}_+\boldsymbol{\alpha}_+$$

or that

$$\mathbf{M}_+^{-1}[\mu(\tau \mid N)]_\tau = [\alpha_1, \alpha_2,..., \alpha_{l(\tau)}, 0,..., 0]'. \tag{7.11}$$

Let η_{ij} denote the ijth entry of $\mathbf{M}_+^{-1}$ and define row operators, for $1 \leqslant j \leqslant n$,

$$R_j^{(\tau)} = \sum_{i=1}^{p} \sum_{l=1}^{n_i} \eta_{j,N_i+l} \cdot \Theta_{i,l-1}; \qquad N_i = \sum_{k=1}^{i-1} n_k\,. \tag{7.12}$$

Then (7.22) is equivalent to

$$\begin{aligned} [R_j^{(\tau)}s]_\tau &= \alpha_j, & 1 \leqslant j \leqslant l(\tau) \\ &= 0, & l(\tau) + 1 \leqslant j \leqslant n. \end{aligned} \tag{7.13}$$

In view of the remarks at the beginning of the section (note that $\mu(t \mid N) = 0$

for $t < t_1$ or $t > t_N$) we have thus established the following generalization of Theorem 3.6 of [1]:

THEOREM 7.1. *$s \in H$ is an Lg-spline interpolating data $\{r_j, 1 \leqslant j \leqslant N\}$ with respect to* EHB $\{\lambda_j, 1 \leqslant j \leqslant N\}$ *on* $[0, T]$ *if and only if*

(a) $L^*Ls(t) = 0$, *when t is not a knot,*

(b) $Ls(t) = 0$, for $t < t_1$ *or* $t > t_N$,

(c) $\lambda_i s = r_i$, $1 \leqslant i \leqslant N$,

(d) $[R_i^{(\tau)} s]_\tau = 0$, *for $l(\tau) + 1 \leqslant i \leqslant n$ if τ is a knot of order $l(\tau)$.* ∎

Practical exploitation of Jerome and Schumaker's theorem has been investigated by various authors (see for example [21–22]).

VIII. Proof of Algorithm (6.20–6.40)

The basic idea is to exploit Theorem 5.2, and the known lumped model (5.12), (6.6–6.13) of the process $y(\cdot)$, to compute the sample function $\hat{y}(t)$ of the l.l.s.e. $\mathscr{P}^{\mathscr{D}_N} y(t)$ corresponding to $\lambda_j y = r_j$, $1 \leqslant j \leqslant N$.

Equations (6.21)–(6.26) are an algorithmic form of (6.6)–(6.13). It relies on the semigroup property of $\phi(\cdot, \cdot)$, namely that

$$\phi(t, \xi)\,\phi(\xi, \tau) = \phi(t, \tau) \qquad \text{all } t, \xi, \tau. \tag{8.1}$$

Hence $\phi(t, \xi)\,\phi(\xi, t) = \phi(t, t) = I_n$ and thus using (6.13):

$$\frac{\partial}{\partial t}\phi(\xi, t) = -\phi(\xi, t)\,A(t). \tag{8.2}$$

Now let $M_j(\cdot)$, $1 \leqslant j \leqslant n-1$, be $j \times n$ matrix-valued functions with ith row $c_i\phi(t_i, \cdot)$. Clearly, then $\mathcal{O} = M_n(t_n)$ and (6.21) and (6.25) are obvious consequences. Furthermore using (8.1), $M_j(t) = M_j(t_j)\,\phi(t_j, t)$ which in view of (7.2) gives us (6.23).

Next, let us define the $n \times n$ matrices, $1 \leqslant j \leqslant n-1$,

$$V_j(t) = [\Delta_1'(t) \mid \cdots \mid \Delta_j'(t) \mid 0]', \tag{8.3}$$

$$\Pi_j(t) = \int_{t_j}^{t} \phi(t, \xi)\,BB'\phi'(t, \xi)\,d\xi, \qquad t_j \leqslant t \leqslant t_{j-1}. \tag{8.4}$$

Thus (6.22) and (6.24) follow readily. Also, an account of (8.1)

$$V_j(t) = \left[\frac{M_j(t_{j+1})}{0}\right]\phi(t_{j+1}, t)$$

and hence from (6.11)–(6.12)

$$Q = \sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} V_i(\xi)\, BB' V_i'(\xi)\, d\xi = \sum_{i=1}^{n-1} \left[\begin{array}{c|c} M_i(t_{i+1})\, \Pi_i(t_{i+1})\, M_i'(t_{i+1}) & 0 \\ \hline 0 & 0 \end{array} \right]. \tag{8.5}$$

Clearly $Q = Q_n$ and the Q_j defined below obey the recursion (6.26):

$$Q_j = \sum_{i=1}^{j-1} \left[\begin{array}{c|c} M_i(t_{i+1})\, \Pi_i(t_{i+1})\, M_i'(t_{i+1}) & 0 \\ \hline 0 & 0 \end{array} \right]. \tag{8.6}$$

For $k \geqslant 1$ let $\mathscr{D}_k = \operatorname{span}\{\lambda_1 y, \ldots, \lambda_k y\}$ and $\mathscr{P}_k$ denote the projection operation onto $\mathscr{D}_k$. First (5.12a) and the linearity of $\mathscr{P}_k$ give $\mathscr{P}_N y(t) = C\mathscr{P}_N x(t)$. Thus (6.20) follows with the notation

$$\begin{aligned} \hat{x}(t \mid k) &= \text{sample value of } \mathscr{P}_k x(t) \text{ corresponding to} \\ &\lambda_j y = r_j, \qquad 1 \leqslant j \leqslant k. \end{aligned} \tag{8.7}$$

We shall also use the following

$$\mathscr{E}_k x(t) \triangleq x(t) - \mathscr{P}_k x(t), \tag{8.8}$$

$$\Pi(t) \triangleq Ex(t)\, x'(t), \qquad \Sigma(t \mid k) \triangleq E[\mathscr{P}_k x(t)]\, [\mathscr{P}_k x(t)]' \tag{8.9}$$

$$P(t \mid k) \triangleq E[\mathscr{E}_k x(t)]\, [\mathscr{E}_k x(t)]'. \tag{8.10}$$

It is clear that $\mathscr{E}_k x(t) \perp \mathscr{P}_k x(t)$ and thus that

$$\Pi(t) = \Sigma(t \mid k) + P(t \mid k). \tag{8.11}$$

As a consequence of (6.15), $\mathscr{P}_n \int_{t_1}^{t_n} \Delta(\xi)\, Bu(\xi)\, d\xi = 0$ and hence from (6.6) $\mathscr{P}_n x(t_n) = \mathcal{O}^{-1}\Lambda$ which readily gives $\hat{x}(t_n \mid n) = \mathcal{O}^{-1}(r_1, r_2, \ldots, r_n)'$ and $\Sigma(t_n \mid n) = \mathcal{O}^{-1}\mathcal{O}^{-T}$. Thus $P(t_n \mid n) = \Pi(t_n) - \mathcal{O}^{-1}\mathcal{O}^{-T}$. In view of (8.11) one then has (6.27–6.28).

The basic idea of the rest of the algorithm is to compute an orthogonal basis $\{\epsilon_j, 1 \leqslant j \leqslant N\}$ for $\mathscr{D}_N$ and then calculate $\mathscr{P}_N x(t)$ as a linear combination of this new basis. In the context of stochastic provesses this Gram–Schmidt procedure has been known as the *Innovations method* [18]. Define

$$\epsilon_j = \lambda_j y - \mathscr{P}_{j-1}(\lambda_j y), \qquad 1 \leqslant j \leqslant N. \tag{8.12}$$

In view of (4.12) and (5.6), for $1 \leqslant j \leqslant n$, $\epsilon_j = \lambda_j y$; and from (6.5), for $j \geqslant n+1$

$$\epsilon_j = \lambda_j y - c_j \mathscr{P}_{j-1} x(t_j) = c_j \mathscr{E}_{j-1} x(t_j). \tag{8.13}$$

Thus if e_j denote the sample values of the ϵ_j we have

$$\begin{aligned} e_j &= r_j\,, & 1 \leqslant j \leqslant n \\ &= r_j - c_j\hat{x}(t_j \mid j-1), & n+1 \leqslant j \leqslant N; \\ R_j^\epsilon = E\epsilon_j^2 &= 1, & 1 \leqslant j \leqslant n \\ &= c_j P(t_j \mid j-1)\, c_j'\,, & j \geqslant n+1 \end{aligned} \tag{8.14}$$

as in (6.29). Now from (8.12) it is clear that

$$E\epsilon_i\epsilon_j = R_i^\epsilon \delta_{ij}; \qquad \mathscr{D}_k = \operatorname{span}\{\epsilon_1\,, \epsilon_2\,, \ldots, \epsilon_k\} \tag{8.15}$$

and thus that for any random variable $w \in Y$,

$$\mathscr{P}_k w = \sum_{j=1}^{k} E\{w\epsilon_j\}\,(R_j^\epsilon)^{-1}\,\epsilon_j = \mathscr{P}_{k-1} w + E\{w\epsilon_k\}\,(R_k)^{-1}\,\epsilon_k; \qquad k \geqslant 1, \tag{8.16}$$

a recursive relation of great importance below.

First (8.16) with $w = x(t_{j+1})$ and $k = j+1$ gives

$$\mathscr{P}_{j+1} x(t_{j+1}) = \mathscr{P}_j x(t_{j+1}) + E\{x(t_{j+1})\,\epsilon_{j+1}\}\,(R_{j+1}^\epsilon)^{-1}\,\epsilon_{j+1} \tag{8.17}$$

from which we have (6.33) with

$$K_{j+1}' = E\{x(t_{j+1})\,\epsilon_{j+1}\} \tag{8.18}$$

Equation (8.17) in turn, upon using (8.10) and (8.13), gives us (6.30). Also, since $\epsilon_{j+1} \perp \mathscr{D}_j$ it is clear that we have

$$\Sigma(t_{j+1} \mid j+1) = \Sigma(t_{j+1} \mid j) + K_{j+1}(R_{j+1}^\epsilon)^{-1}\,K_{j+1}'\,. \tag{8.19}$$

This together with (8.11) gives (6.34). Next, from (5.12) and (6.13) one has

$$x(t) = \phi(t, t_j)\,x(t_j) + \int_{t_j}^{t} \phi(t, \xi)\,Bu(\xi)\,d\xi \tag{8.20}$$

which, for $t \geqslant t_j \geqslant t_n$, gives (recall that $Eu(\xi)\,x(t_n) = 0$, $\xi \geqslant t_n$)

$$P_j x(t) = \phi(t, t_j)\,P_j x(t_j) \tag{8.21}$$

of which (6.31) is a ready consequence. Also we then have $\Sigma(t \mid j) = \phi(t, t_j)\,\Sigma(t_j \mid j)\,\phi'(t, t_j)$ and from (8.20),

$$\Pi(t) = \phi(t, t_j)\,\Pi(t_j)\,\phi'(t, t_j) + \int_{t_1}^{t} \phi(t, \xi)\,BB'\phi'(t, \xi)\,d\xi,$$

the two of which together with (8.11) give (6.32).

Furthermore from (8.20)–(8.21)

$$\mathscr{E}_j x(t) = \phi(t, t_j)\, \mathscr{E}_j x(t_j) + \int_{t_j}^{t} \phi(t, \xi)\, Bu(\xi)\, d\xi.$$

Set $t = t_{j+1}$ and for $\mathscr{P}_j x(t_j)$ substitute from (8.17). Thus

$$\mathscr{E}_j x(t_{j+1}) = \Psi_{j+1,j} \mathscr{E}_{j-1} x(t_j) + \int_{t_j}^{t_{j+1}} \phi(t_{j+1}, \xi)\, Bu(\xi)\, d\xi \tag{8.22}$$

where

$$\Psi_{j+1,j} = \phi(t_{j+1}, t_j)\{I_p - K_j(R_j^{\epsilon})^{-1} c_j\}. \tag{8.23}$$

Next for $t \geqslant t_N$

$$x(t) = \phi(t, t_N)\, x(t_N) + \int_{t_N}^{t} \phi(t, \xi)\, Bu(\xi)\, d\xi.$$

Thus in view of (6.8)

$$\mathscr{P}_N x(t) = \phi(t, t_N)\, \mathscr{P}_N x(t_N) \qquad \text{for } t \geqslant t_N$$

or $\hat{x}(t \mid N) = \phi(t, t_N)\, \hat{x}(t_N \mid N)$ which gives (6.35) for the case of $t \geqslant t_N$. The case of $t \leqslant t_1$ is similar.

Let $t_n \leqslant t \leqslant t_N$. Suppose now that $t_{j-1} \leqslant t \leqslant t_j$ then

$$x(t) = \phi(t, t_{j-1})\, x(t_{j-1}) + \int_{t_{j-1}}^{t} \phi(t, \xi)\, Bu(\xi)\, d\xi$$

or

$$\mathscr{P}_N x(t) = \phi(t, t_{j-1})\, \mathscr{P}_N x(t_{j-1}) + \mathscr{P}_N \int_{t_{j-1}}^{t} \phi(t, \xi)\, Bu(\xi)\, d\xi. \tag{8.24}$$

But it is clear that since $E\{u(\xi)\, x'(\tau)\} = 0$ for $\xi > \tau$ we have $E\{u(\xi)\, \epsilon_l\} = 0$ for $l \leqslant j - 1$. Thus

$$\mathscr{P}_N \int_{t_{j-1}}^{t} \phi(t, \xi)\, Bu(\xi)\, d\xi = \sum_{l=j}^{N} \int_{t_{j-1}}^{t} \phi(t, \xi)\, BE\{u(\xi)\, \epsilon_l\}\, (R_l^{\epsilon})^{-1}\, \epsilon_l\, d\xi. \tag{8.25}$$

But using (8.13):

$$E\{u(\xi)\, \epsilon_l\} = E\{u(\xi)\, \mathscr{E}_{l-1} x'(t_l)\}\, c_l'. \tag{8.26}$$

which upon repeatedly using (8.22) gives

$$\begin{aligned} E\{u(\xi)\, \epsilon_l\} &= \left[E\{u(\xi)\, \mathscr{E}_{j-2} x'(t_{j-1})\}\, \Psi'_{j,j-1} + \int_{t_{j-1}}^{t_j} E\{u(\xi)\, u'(\theta)\}\, B'\phi'(t_j, \theta)\, d\theta\right] \Psi'_{l,j} c_l' \\ &= B'\phi'(t_j, \xi)\, \Psi'_{l,j} c_l'; \qquad l \geqslant j; \quad t_{j-1} < \xi < t_j. \end{aligned} \tag{8.27}$$

Thus from (8.25)–(8.27)

$$\hat{x}(t \mid N) = \phi(t, t_{j-1})\, \hat{x}(t_{j-1} \mid N) + \int_{t_{j-1}}^{t} \phi(t, \xi)\, BB'\phi'(t_j, \xi)\, d\xi \cdot \sum_{l=j}^{N} \Psi'_{l,j} c'_l (R_l^{\epsilon})^{-1} e_l . \tag{8.28}$$

which gives

$$\frac{d}{dt}\hat{x}(t \mid N) = A(t)\, \hat{x}(t \mid N) + BB' \left\{ \phi'(t_j, t) \sum_{l=j}^{N} \Psi'_{l,j} c'_l (R_l^{\epsilon})^{-1} e_l \right\} .$$

This in turn gives (6.35) with

$$\mu(t \mid N) = \mu_j(t \mid N) = \phi'(t_j, t) \sum_{l=j}^{N} \Psi'_{l,j} c'_l (R_l^{\epsilon})^{-1} e_l; \qquad t_{j-1} \leqslant t \leqslant t_j . \tag{8.29}$$

Equations (6.37), (6.40) are then immediate. Note, also that

$$\mu_j(t_j \mid N) = \sum_{l=j+1}^{N} \Psi'_{l,j} c'_l (R_l^{\epsilon})^{-1} e_l + c'_j (R_j^{\epsilon})^{-1} \epsilon_j$$

which upon using (8.29) gives (6.39a).

The case of $t_1 < t < t_n$ is a little more difficult. First of all as in (8.28), now

$$\hat{x}(t \mid N) = \phi(t, t_n)\, \hat{x}(t_n \mid N) + \sum_{l=n+1}^{N} \int_{t_n}^{t} \phi(t, \xi)\, BE\{u(\xi)\, \epsilon_l\}\, (R_l^{\epsilon})^{-1} \epsilon_l\, d\xi. \tag{8.30}$$

But using (8.13) and (8.22) [note also the whiteness of $u(\cdot)$, (5.11)]

$$E\{u(\xi)\, \epsilon_l\} = E\{u(\xi)\, x'(t_{n+1})\}\, \Psi'_{l,n+1} c'_l , \qquad l \geqslant n+1.$$

However $x(t_{n+1}) = \phi(t_{n+1}, t_n)\, x(t_n) + \int_{t_n}^{t_{n+1}} \phi(t_{n+1}, \xi)\, Bu(\xi)\, d\xi$. Thus (use (6.8))

$$\begin{aligned} E\{u(\xi)\, \epsilon_l\} &= E\{u(\xi)\, x'(t_n)\}\, \phi'(t_{n+1}, t_n)\, \Psi'_{l,n+1} c'_l \\ &= B'\Delta'(\xi)\, \mathcal{O}^{-T} \phi'(t_{n+1}, t_n)\, \Psi'_{l,n+1} c'_l , \qquad l \geqslant n+1. \end{aligned} \tag{8.31}$$

Thus from (8.29)–(8.31):

$$\frac{d}{dt}\hat{x}(t \mid N) = A(t)\, \hat{x}(t \mid N) + BB'\Delta'(t)\, \mathcal{O}^{-T} \mu_{n+1}(t_n \mid N)$$

which is (6.35) with

$$\mu(t \mid N) = \Delta'(t)\, \mathcal{O}^{-T} \mu_{n+1}(t_n \mid N), \qquad t_1 \leqslant t \leqslant t_n .$$

Then (6.38b)–(6.39) are easily shown using (6.10) and (6.12).

IX. Concluding Remarks

This paper has been aimed at extending logically and completely the existent theory of Lg-splines to the situation of simultaneous interpolation of several functions. In later papers in the series we shall examine the same extension for smoothing splines and ARMA splines. In particular, the latter brings with it an increasing complexity as far as the various aspects considered in Appendix I are concerned.

APPENDIX I: The Structure of $Ly = u$

We have defined in (2.2) the $p \times p$ matrix $L(D)$ of ordinary differential operators $L_{ij}(D)$ subject to the restriction (2.3).

Given a $p \times p$ matrix $M(D)$ of polynomials let p_j, $1 \leqslant j \leqslant p$, denote the degree of the highest degree term in column j of $M(D)$. Now, let us define Γ_M as the matrix obtained as follows. Place in each position ij of Γ_M the coefficient of D^{p_j} in the ijth entry of $M(D)$. Matrix $M(D)$ is said to be *column proper* [20] if Γ_M is nonsingular. Now, restriction (2.3) ensures that for $L(D)$, we have $p_j = n_j$, $1 \leqslant j \leqslant p$, and that Γ_L is a $p \times p$ identity matrix and hence that $L(D)$ is column proper. This is central to the rest of the development.

Using various, by now well-known, results on the canonical forms of linear systems (see for example [20]) it is easily shown that the system of p differential equations

$$Ly = u \tag{A.1}$$

with $y \in H$ and $u \in \mathscr{L}_2{}^p$ has the state-space representation of equations (5.12). This can be verified by direct expansion of (5.12). Now we establish Lemma 3.1.

Proof of Lemma 3.1. Note that $y \in N_L$ if and only if $Ly = 0$. In terms of (5.12), $N_L = \{y \in H: y(t) = Cx(t), dx(t)/dt = A(t)\, x(t)\}$. It is easily seen then that, picking any $t_0 \in W$, $y \in N_L$ can be written as $y(t) = C\phi(t, t_0)\, x_0$ where $x_0 = x(t_0)$ is an arbitrary $n \times 1$ real vector. In other words

$$N_L = \operatorname{span}\{\text{columns of } C\phi(\cdot, t_0)\}. \tag{A.2}$$

However we can show that these n columns are independent in H and hence that $\dim N_L = n$. To do so it is enough to show that $y(\cdot) = C\phi(\cdot, t_0)\, x_0 = 0$ implies $x_0 = 0$. But using (5.12), with $u(\cdot) \equiv 0$, it is easily shown that $y(\cdot) \equiv 0$ implies $\{y_j^{(k)}(t_0) = 0, 1 \leqslant j \leqslant p, 0 \leqslant k \leqslant n_j - 1\}$ and thus in view of (5.12g) that $x_0 = x_0(t_0) = 0$, which completes the proof. ∎

Strictly speaking, restriction (2.3) can be eliminated. All that is needed is that $L(D)$ be a column proper polynomial matrix. Then Lemma 3.1 and all the subsequent results continue to hold. However, $A(\cdot)$, B, and C become more complex in structure. One can then, in fact, reduce the equations $L(D)\,y = u$, by forming suitable linear combinations, to a new form $\bar{L}(D)\,y = \bar{u} = \bar{M}u$, where $\bar{M}$ is an invertible $p \times p$ matrix of constants such that $\bar{L}(D) = \bar{M}L(D)$ satisfies the condition $\Gamma_{\bar{L}} = I_p$; then $\bar{L}(D)\,y = \bar{M}u$ gives rise to suitable $A(\cdot)$, B, C matrices with minor changes in B and C.

A more difficult situation arises when $L(D)$ is not column proper. Then a suitable row operation procedure [20] is needed to reduce it to a column proper form $\bar{L}(D)$. But then the column degrees of $\bar{L}(D)$ obey $p_j \leqslant n_j$, and consequently $n = \dim N_L \leqslant \sum_{j=1}^{p} n_j$; with this modification the rest of the results continue to hold, modulo certain changes of detail.

References

1. J. W. Jerome and L. L. Schumaker, On Lg-splines, *J. Approximation Theory* **2** (1969), 29–49.
2. H. L. Weinert and G. S. Sidhu, On uniqueness conditions for optimal curve fitting, *J. Optimization Theory Appl.* **2** (October 1977), 211–216.
3. A. N. Netravali and R. J. P. de Figueiredo, On a class of minimum energy controls related to spline functions, *IEEE Trans. Automatic Control. AC*-**21** (1976), 725–727.
4. H. L. Weinert and T. Kailath, A spline-theoretic approach to minimum energy control, *IEEE Trans. Automatic Control AC*-**21** (1976), 391–393.
5. H. L. Weinert and G. S. Sidhu, "A Spline Theoretic Approach to Minimum Energy Control—Part II: Systems with Numerator Dynamics and ARMA Splines," Tech. Rpt. 75-11, Dept. of Electrical Engineering, The Johns Hopkins University, Baltimore, Maryland, 1975.
6. R. J. P. de Figueiredo, "LMg-Splines," Report 75-10, Dept. of Electrical Engineering, Rice University, Houston, Texas, 1975.
7. D. G. Luenberger, "Optimization by Vector Space Methods," Wiley, New York, 1969.
8. N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.* **68** (1950), 337–404.
9. J. Mercer, Functions of positive and negative type and their connection with the theory of integral equations, *Philos. Trans. Roy. Soc. London Ser. A* **209** (1909), 415–446.
10. G. S. Sidhu and H. L. Weinert, Dynamic recursive algorithms for Lg-spline interpolation of EHB data, *Appl. Math. and Computation*, in press.
11. E. Parzen, "Time Series Analysis Papers," p. 267, Holden–Day, San Francisco, 1967.
12. M. Loève, Fonctions aleàtoires du second ordre, *in* "Processus stochastiques et mouvement Brownien" (P. Lévy, Ed.), 2nd ed., pp. 367–420, Gauthier–Villars, Paris, 1965.
13. G. Kimeldorf and G. Wahba, Spline functions and stochastic processes, *Sankhyā Ser. A* **132** (1970), 173–180.
14. G. Kimeldorf and G. Wahba, A correspondence between Bayesian estimation on stochastic processes and smoothing by splines, *Ann. Math. Statist.* **41** (1970), 495–502.
15. H. L. Weinert, "A Reproducing Kernel Hibert Space Approach to Spline Problems with Applications in Estimation and Control," Ph.D. thesis, Dept. of Electrical Engineering, Stanford University, Stanford, Calif., 1972.

16. H. L. WEINERT AND T. KAILATH, Stochastic interpretations and recursive algorithms for spline functions, *Ann. of Statist.* **2** (1974), 787–794.
17. H. L. WEINERT AND G. S. SIDHU, A stochastic framework for recursive computation of spline functions—Part I: Interpolating splines, *IEEE Trans. Information Theory*, **IT-24** (1978), 45–50.
18. T. KAILATH, The innovations approach to detection and estimation theory, *Proc IEEE* **58** (May 1970), 680–695.
19. H. L. WEINERT AND G. S. SIDHU, "Stochastic Error Analysis of Spline Interpolation," JHU-EE Report 77-4, Dept. of Electrical Engineering, The Johns Hopkins University, Baltimore, Maryland, 1977.
20. W. A. WOLOVICH, "Linear Multivariable Systems," Springer-Verlag, New York/Berlin, 1974.
21. C. CARASSO AND P. J. LAURENT, On the numerical construction and practical use of interpolating spline functions, "Proc. IFIP Congress Information Processing 68 (Edinburgh 1968)," Vol. 1, pp. 86–89, *MR* **40**, No. 8219.
22. M. J. MUNTEANU AND L. L. SCHUMAKER, On a method of Carasso and Laurent for constructing interpolating splines, *Math. Comp.* **27**, No. 122 (April 1973), 317–325.